



# THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### A new test of legibility

**Citation for published version:**

Van Rossum, M 1997, 'A new test of legibility' *Quaerendo*, vol 27, no. 2, pp. 141-147. DOI: 10.1163/157006997X00222

**Digital Object Identifier (DOI):**

[10.1163/157006997X00222](https://doi.org/10.1163/157006997X00222)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

*Quaerendo*

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



MARK VAN ROSSUM

## A new test of legibility

This article describes a new method of measuring the legibility of text based on the reading process and the properties of the human eye. The new test is both more accurate and more objective than traditional tests. It has been used to test the legibility of a variety of typefaces including Gulliver, a new face designed by Dutch type designer Gerard Unger. The new legibility test also clarifies the importance of x-height and the function of serifs.

### INTRODUCTION

Most people find reading quite an easy task. In fact, however, it is a singularly complicated process. Apparently without effort, a particular piece of text is selected from all we see, and characters are recognized. The complexity of the reading process becomes clear when we consider the enormous range of conditions under which we can read: different typefaces and handwriting styles, different sizes, text set horizontally and vertically, text seen close to and at a distance, at an angle, upside-down and so on.

Despite the fact that one can adjust to many reading circumstances, some typefaces are more readily or more comfortably read than others; yet it is often unclear why. The most obvious test is to set a sample text in various typefaces and then proceed to measure how fast a large number of subjects can read it. Tests of this kind have been carried out many times, investigating all sorts of factors such as type size, line spacing and line length. The results of these studies may be summarized as follows: (1) We do not read letter by letter, but recognize a word or group of letters all at once. (2) Legibility is determined not so much by the overall size of the letters but by their x-height (the x-height is the height of the letter x, and thus equals the overall height of the letter less the length of any ascender and/or descender). The x-height and overall height are of course connected, but a type designer has a certain amount of freedom in choosing their proportions. It is a remarkable fact that typefaces with different overall heights but the same x-height *look* the same size.

These studies have also looked at different sorts of typeface, but little difference has been found in the legibility of the common faces. The only clear difference is that text set in italic faces, or exclusively in capitals, is less legible. Nor has any clear difference been found in the legibility of seriffed as opposed to sanserif

faces. A good overview of almost a century of legibility research will be found in Spencer (1968).

Clearly, then, our eyes and brain are capable of reading common typefaces accurately and fast. However, it is still very important to know which typeface is the most legible, even if the differences are slight. The point is that more legible typefaces can be used in smaller sizes – and setting a text a few per cent smaller can yield massive savings in printing and paper costs. Thus there is a need for a test which will be both more accurate than traditional tests and, if possible, easier to perform.

The legibility test presented here has been designed to allow accurate straight comparisons between different typefaces. There is no need to have dozens of subjects. What prompted the research was the arrival of Gulliver, a new typeface which has been designed to be more legible than the common typefaces, meaning that it can be set smaller. Figure 1 is an example of Gulliver. The test not only shows that Gulliver is more legible, it also measures the extent of the saving achieved.

## Rijksinstituut voor Volksgezondheid en Milieuhygiëne

1. The Gulliver typeface. Gulliver is characterized by large counters and a relatively large x-height.

### HOW DO WE READ?

Before describing the new method, it will be useful first to study the reading process. As we read, the eye does not glide smoothly along the line but makes a succession of small jumps. Every time it does so, a block of some 5 to 20 characters is recorded, after which the eye moves on to fixate on the next block (Rayner & Polatsek, 1987). These fixations have important consequences. We know that the sharpness of the image produced by the eye diminishes rapidly with increasing distance from the centre of the visual field (Adler, 1981, and Kroner & Kaplan, 1995). This means that most of the characters that we see during a fixation are not particularly sharp (see figure 2). Free-standing letters on the edge of the visual field may still be recognizable, but groups of letters in this area soon become unrecognizable (Bouma, 1970). This is easily demonstrated. Close one eye and focus on a single spot in the text, then try to read the surrounding words. It will be found that only words very close to the point of focus can be read. This also explains why type that is too small is so unpleasant to read. To most people a much smaller type than the common ten or twelve-point still seems reasonably legible – legible, that is, when it is in the centre of the visual field. Outside the centre, however, it will be illegible, so the eyes have to make many small steps, which is why reading small print is slower

and more uncomfortable. (If the lines are very short, however, as in a telephone book or timetable, small type works well.) If type is to be legible, readers have to be able to absorb as many characters as possible at each fixation. This means that letters have to stay legible even at the edges of our visual field. We therefore define legibility as follows: *Legibility is proportional to the average number of letters in a line that can be recognized while keeping the eye fixed.*



2. The acuity of the eye declines sharply outside the centre of the visual field. The figure shows how we simulate this. The more letters are still recognizable, the more legible is the typeface.

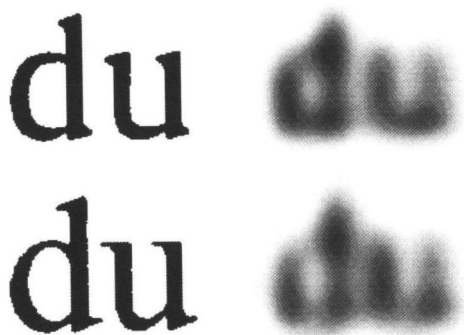
#### THE NEW TEST OF LEGIBILITY

The new test determines whether words away from the centre of the visual field, where the acuity of the eye diminishes, continue to be legible. We do not know exactly why and by what mechanism the acuity of vision declines at the edges, but it is believed the eyes are not the whole story: the brain too has a part to play. We therefore looked for a simple solution: we simulate the decline in sharpness by making the image of a text increasingly fuzzy by deliberately blurring it. This is done by convoluting the image with a Gaussian function, so that each pixel is smeared. The Gaussian function is a bell-shaped curve. When this is combined with the printed letters it adds fluid transitions and produces the smearing effect. This sounds complicated, but the result is roughly comparable with an illustration that has been photographed out of focus. Figure 3 shows the sample text after this operation. By increasing the width of the Gaussian function, the text is then made increasingly fuzzy and the width of the Gaussian function at which it is no longer legible is determined. This is then termed the *maximum blurring width*. Our definition of legibility now becomes: *Legibility is the product of the mean maximum blurring width and the number of letters per unit length*. Note that when this definition is used with larger sizes of the same typeface the maximum blurring width increases but there are also fewer letters to the line so that legibility remains the same.

This test measures only the differences between individual typefaces. Other aspects that only come into play when types are actually used, such as size, colour, contrast, line length etc., must be measured with more conventional tests. This is an advantage because such factors are often subjective.

It is interesting first to examine some general insights which followed from this new method. To begin with, the method confirms the importance of

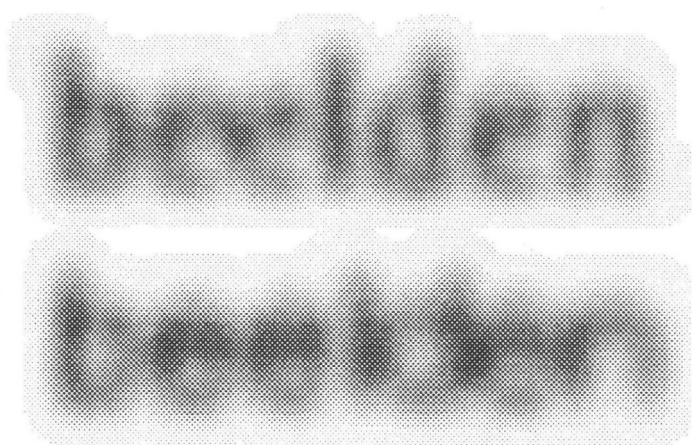
x-height. Types with the same x-height have roughly the same maximum blurring width. This explains why types with the same x-height give the visual impression of being the same size. Second, after blurring, letters such as c and e, or i and l, are easily confused; k and w, by contrast, are robust and are not easily confused with other letters. This much we already knew from conventional legibility research. However, it turns out to be very important not to look at individual, free-standing letters, but to look instead at the letters in the context of the words in which they are used. The blurring procedure not only blurs the individual letters, it also causes adjacent letters to run into each other. Combinations such as nn, ij or lt are very susceptible to this. Combinations such as oo or vi are relatively unaffected by mutual blurring because the main parts of the letters are further apart or because they are better able to preserve their own character. The letter d, for example, was found by several researchers to be either a very legible or even the most legible letter of all. But using our new method to look at combinations like du or dl, the d soon becomes unrecognizable (see figure 3). So investigating individual letters is not particularly helpful.



3. An example of the blurring test. Although the d is often cited as an example of a very clear letter, in some combinations it quickly becomes illegible. Top Gulliver, bottom Times.

As a third point, the test also reveals how useful serifs are: when blurring is introduced, the serifs ensure that the definition of the verticals remains intact. The ascenders in the sanserif Helvetica soon disappear in the 'mist' (see figure 4). In some letter combinations, on the other hand, Helvetica benefits from its clean style. Nevertheless, as we shall see below, types with serifs do prove generally more legible.

These insights can only be found using our test, and are valuable for type design.



4. The use of serifs. The ascenders and descenders of letters with serifs are still clearer than those of sanserif types after blurring. Even so, in some words Helvetica is more legible because of its cleaner forms. Top Gulliver, bottom Helvetica. The word shown is 'beelden.'

#### COMPARISON OF VARIOUS TYPEFACES

The legibility of four different typefaces was measured using an arbitrarily chosen Dutch text which was scanned in at high resolution (600 dpi) and stored in a computer. The x-heights of all four faces were the same at approx. 2.5 mm (roughly 16 pt. depending on the typeface). The texts were then blurred using a Gaussian function with a width of between 0.8 and 1.1 mm. For practical purposes the text was first broken down into separate words or word-parts of 2-3 letters. Thus the word 'op' was not broken down, but 'bij-schrift' was split into 'bij-schr-ift.' The assessment of whether a word or word-part was or was not recognizable was made visually, using a small window which revealed only a few letters of the word and thus avoided the natural tendency to deduce the word-part from the rest of the word. In all, 40 arbitrarily chosen Dutch words were used, each being split into a total of 92 word-parts each of which consisted of several letters. The researcher then determined when each word-part was still just legible, and in this way the maximum blurring width was established. The maximum blurring width was averaged over the 92 word-parts.

The familiar Times (seriffed, Monotype) was compared with Helvetica (sanserif, Linotype-Hell), Argo (sanserif, DTL) and Gulliver (seriffed). Times and Helvetica were chosen because both types are extremely common – so much so, indeed, that they are probably the most widely used typefaces in the world. Argo is a sanserif face, but in contrast to Helvetica the thickness of the lines varies. Gulliver was developed specifically for the purpose of saving space. This has been achieved

by using large counters and a relatively large x-height. The difference between Gulliver and Times can be readily appreciated from the left-hand half of figure 3. At the same x-height, Gulliver takes up less space than the other faces and is accordingly more legible. Times serves as a reference point, its legibility being taken to be 100%. The results are shown in table 1. The maximum blurring width for Helvetica was 5.4% smaller and for Gulliver 0.5% smaller than that for Times. In our definition of legibility the amount of horizontal space a text occupies is crucial. The text in Helvetica took up 3.5% less space than in Times; the same text set in Gulliver took up 6.2% less. This means that the overall legibility of Helvetica is 1.9% less than that of Times, that of Argo 3.2% less. Gulliver, on the other hand, has legibility 5.7% *better* than that of Times. To put it another way, at the same level of legibility the text in Gulliver takes up 5.7% less space on the page.

typeface	maximum blurring width	length of text	legibility
Helvetica	-5.4%	-3.5 %	-1.9%
Argo	-3.0%	+0.2%	-3.2%
Gulliver	-0.5%	-6.2%	+5.7%

Table 1: Results for three typefaces relative to Times. Gulliver is 5.7% more legible than Times. The margin of error in legibility is  $\pm 2\%$ , i.e. the legibility of Helvetica is between 0% and 4% less legible than Times.

We also compared the book (a slightly heavier version than normal), semi-bold and bold versions of Gulliver against the normal version. All versions had roughly the same blurring width. Of course, bold types take up more space, but in the end only the boldest version is significantly less legible (some 4%).

The leading of all texts was the same. However, using types of smaller overall height means that it is possible to place the lines more closely together, which in turn means a larger overall gain in space: how much larger is difficult to determine with this test, but several per cent of extra space saved would certainly be possible. With sanserif faces the lack of serifs means that the base line is less well defined, which is why sanserifs generally need rather greater leading.

We have already observed that the blurring width and hence legibility depends principally on x-height, rather than on overall height. This means that it is actually possible to make a letter smaller by increasing its x-height, as has been done with Gulliver. Obviously, there are limits to how far this can be taken. The new test can reveal where the limit lies.

The test used here is a first step on the way to a wholly objective test of the legibility of different typefaces. In the future it may be possible for computers using neural networks to do the 'reading' of blurred texts and thus determine the legibility of different faces. That will make the test truly objective. It is also possible to use the test during the design process of new typefaces. It can be implemented reasonably well on the current generation of PCs so that designers can conduct their own tests – though naturally all sorts of factors in the design, such as field of application, aesthetic considerations and consistency, will still depend on the skill of the designer. But because we can now see precisely which blurred letters are still legible and which are not, stumbling-blocks in the design can now be identified at an early stage.

#### BIBLIOGRAPHY

- Spencer, H., *The Visible Word* (London: Royal College of Art 1968)  
Tinker, M.A., *Legibility of Print* (Ames, Iowa: Iowa State University Press 1963)  
Bouma, H., 'Interaction effects in parafoveal letter recognition', in: *Nature*, 226 (1970), p. 177  
Adler, F.H., *Adler's Physiology of the Eye*, ed. R.A. Moses (St. Louis: The C.V. Mosby Company 1981)  
Croner, L.J. & Kaplan, E., 'Receptive fields of P and M ganglion cells across the primate retina', in: *Vision Research*, 35 (1995), pp. 7-24  
Rayner, K., & Polatsek, A., 'Eye movements in reading: A tutorial review', in: *Attention and Performance XII: The Psychology of Reading*, ed. M. Colthaert (Hove 1987)